Introduction to Security and Privacy on the Blockchain

Harry Halpin
Inria
Paris, France
harry.halpin@inria.fr

Marta Piekarska
Blockstream
Montreal, Quebec
marta@blocksteam.io

***Abstract*—The blockchain has fueled one of the most enthusiastic bursts of activity in applied cryptography in years, but outstanding problems in security and privacy research must be solved for blockchain technologies to go beyond the hype and reach their full potential. At the first IEEE Privacy and Security on the Blockchain Workshop (IEEE S&B), we presented peer-reviewed papers bringing together academia and industry to analyze problems ranging from deploying newer cryptographic primitives on Bitcoin to enabling use-cases like privacy-preserving file storage. We overview not only the larger problems the workshop has set out to tackle, but also outstanding unsolved issues that will require further co-operation between academia and the blockchain community.**

I. Blockchain Research Challenges

A blockchain is simply a cryptographically verifiable list of data. One of the reasons for the enthusiasm around the blockchain is that databases do not have any cryptographic guarantees of integrity, guarantees that are necessary for any database operating in an adversarial environment. If the field of systems security and privacy-enhancing technologies has learned one lesson since the Snowden revelations, it is that all databases are likely operating in an adversarial environment. Therefore, some of the "hype" around blockchains is for good reason: For the first time in decades, the venerable database itself may be replaced by blockchains. However, there is more to blockchains than just data integrity. As exemplified by Bitcoin, the primary advantage of blockchain technologies is that the data itself can be decentralized. A distributed public ledger built with a blockchain where all users have the same data, which is necessary for high-value use-cases such as currency, is clearly privacy-invasive for many use-cases. Security and privacy on the blockchain is an emerging field that is dire need of further research.

While the first proposals for cryptographic hash functions and distributed systems date from at least the 1970s, these powerful concepts were brought together in the anonymous whitepaper of Satoshi Nakamoto that detailed "Bitcoin: A peer-to-peer electronic cash system" in 2008. Bitcoin was not purely academic, but implemented in the wild: The open source code of Bitcoin made money transfers without a bank acting as a trusted third party possible for millions of users, and its distributed design gave Bitcoin the properties of a permissionless network with censorship-resistance. However, Bit-coin's design still struggles to ensure some measure of anonymity, despite the fact that most of its users believe it provides anonymous payments [3]. Research has been made even more difficult as the privacy and security properties of Bitcoin were never formally stated by Nakamoto in a provable manner, and so these properties have only recently begun to be formalized [1]. As the properties and underlying formal foundations of blockchain technologies are still under debate, in practical terms new blockchains with security and privacy claims seem to be coming out increasingly rapidly. Now more than ever, the academic community should engage to separate the wheat from the chaff, the blockchain scams from the substantial contributions.

The study of security and privacy on the blockchain is growing rapidly but this is the first academic workshop devoted primarily to the topic. By organizing the workshop, we hope to further involve the security and privacy research community in the more ad-hoc and informal world of blockchain programmers in order to tackle the privacy and security properties of both Bitcoin and its underlying blockchain technology. A mutually beneficial cycle between real-world code and academic research is needed, where academia would gain novel problems to solve and developers would be informed of solutions to issues they would otherwise find insoluble.

II. Improvements to Core Cryptographic Primitives

Blockchain technologies are now primarily industry-driven, lending the field a slightly different angle than other areas of research: With Bitcoin and variants being developed by practitioners rather than cryptographers, the trust tends to be put not in formal proofs and properties but in practical resistance to attacks based on common knowledge and experience by practitioners. What follows from this design is a constant fluidity of proposed solutions as well as a lack of common, unified design choices and criteria. Thus, there is a plethora of solutions that each claim to be the best solution. It will be interesting to observe how the field evolves and what ends up being chosen as the golden standard, but any promise of stability is in the distant future, especially in more experimental blockchains such as side-chains.

The natural starting point of blockchain research around the security and privacy properties of the system would

be creating a common vocabulary. Once this has been done, these informal definitions can be given rigorous definitions that would allow systems to be concretely tested in terms of whether or not a particular property was fulfilled. Goldwasser and Micali's "Probabilistic Encryption" [2] formalized the notions of security in terms of a rigorous definition of semantic security. Researchers into blockchain technologies struggle even with the term itself: Is the term "blockchain" singular or plural? Does it refer to Bitcoin or to any system with a decentralized ledger such as Ethereum? Would a ledger operating without mining, such as Hyperledger, also count? In detail, does "the blockchain" refer to to any linked list built out of pointers that are cryptographic hashes, or does it refer to particular design choices done by Bitcoin? Potentially better, more efficient data structures could be proposed to replace the blockchain itself while still fulfilling its role as a distributed public ledger, and other designs would ditch the role of a public ledger in favor of better privacy properties. More work on formalizing terms such as the blockchain, side-chains, public ledger, proof of work, proof of stake, and other components and design choices need to be made, and so the contributed papers themselves vary in their use of terminology around blockchains.

One of the parts of the Bitcoin blockchain that has attracted considerable critique is the use of "proof of work" in order to prevent sybil attacks, where a malicious attacker can flood a decentralized network with users that are secretly controlled by the attacker. Although the use of hashing as proof-of-work allows new participants to join the blockchain in a 'permissionless' fashion seems to have worked so far, a fundamental re-thinking of how to prevent sybil attacks is given in the full paper "Proof-of-Personhood: Redemocratizing Permissionless Cryptocurrencies." While it does seem some sort of high integrity distributed data store is here to stay, such techniques that allow new identifies to participate in a permissionless blockchain without allowing censorship may be necessary in order to let cryptocurrencies scale.

Given the ad-hoc nature of many of the core design choices of Bitcoin, one important consideration is whether or not it is possible to isolate the cryptographic primitives to allow secure and painless upgrades to the Bitcoin blockchain. For example, concerns have been raised over the ECDSA signature scheme being based on an efficient but idiosyncratic *secp256k1* curve. Thus, there is a proposal to adopt the high-speed constant-time EdDSA signature scheme, but this scheme would not allow for deterministic generation of new keys. The paper "BIP32-Ed25519: Hierarchical Deterministic Keys over a Non-linear Keyspace" demonstrates exactly how Ed25519 keys can be used to be compatible with software libraries built for EdDSA and Bitcoin key generation. This work is well complemented by a short paper called "Long-term public blockchain: Resilience against Compromise of Underlying Cryptography" on methods of transitioning to new cryptographic primitives in a general sense, proposing a solution that involves archiving older blocks in a uniform manner and verifying the transition to newer blocks using a new hash function or digital signature scheme. A proposal for a seamless upgrade of cryptographic primitives would be required in any change to the core primitives of the Bitcoin blockchain, including post-quantum primitives in the case of digital signatures.

III. Privacy and Anonymity on the Blockchain

Bitcoin has not been immune from confusion over pseudonymity, privacy and anonymity. Bitcoin offers only pseudonymous transactions. As Bitcoin is a shared global public ledger, pseudonyms may be de-anonymized by determining patterns of usage in the blockchain. This is in contrast to previous well-studied and even deployed systems based on centralized Chaumian e-cash that have well-defined privacy properties, such as enforcing privacy between senders and recipients of transfers, given by well-known primitives such as blind signatures. Yet it seems a public ledger is needed to prevents double-spending attacks, so any user can read and possibly write new information to let some information become public knowledge. Therefore, large changes are needed to existing blockchain technologies in order to preserve privacy. We have seen two approaches to the problem. One is to add anonymization (or at least, some greater privacy) to the existing blockchain by techniques such as Confidential Transfers. The other possible method is to create new blockchains that are incompatible with Bitcoin, such as Zerocash that offer guarantees around anonymity built-in by the use of new primitives in their blockchain, in particular zero-knowledge "succinct non-interactive argument of Knowledge" (SNARKS)[4]. The current design of Zerocash has no audibility, a property possessed by Chaumian e-cash as deployed by DigiCash. The paper "Auditable Zerocoin" allows Zerocoin to be audited, given that auditing may be needed to prevent charges such as those of being a money transmitter; charges that destroyed pre-Bitcoin electronic currencies such as eGold.

A number of companies claim to increase privacy through some kind of mixing schemes, where Bitcoin transactions from different users were mixed together. This however meant relying on a third party, which not always proved to be secure or beneficial[3]. A number of more well thought-out generic techniques, such as Coinjoin (joining multiple payments) and Confidential Transactions have been created that have better anonymity guarantees, and anonymously created software such as Mimblewimble claims to achieve very strong anonymity properties but breaks compatibility with Bitcoin. However, we have not observed a wide scale adoption of any anonymity solution for Bitcoin, which reduces the anonymity set and thus the effectiveness of the proposed solutions. The first large-scale study of anonymization techniques on the Bitcoin blockchain is pre-

sented in "Anonymous Alone? Measuring Bitcoin's Second-Generation Anonymization Technique." We expect future work to continue to study de-anonymization of blockchain records even when new anonymizing techniques are used, due to common patterns in usage such as regular payments, as explored in the short paper "Conditions of Full Disclosure: The Blockchain Remuneration Model." Whether or not these techniques can withstand the advances of machine-learning is of immense interest to both the research community and ordinary users who want to use Bitcoin with a measure of privacy.

Lastly, increased anonymity may allow whole new uses of blockchain technologies. One of the more classical yet exciting use-cases would be the sharing of files in a more anonymous fashion than allowed by current P2P networks based on Bittorrent. Based on linkable ring signatures, financially incentives and privacy-preserving file-sharing is presented in "Design of a Privacy-Preserving Decentralized File Storage with Financial Incentives." While anonymity was sacrificed in the original design of Bitcoin in order to prevent double-spending, we expect this to be a fertile area for future applications that attempt to re-balance Bitcoin's original design choices towards privacy in a wide variety of decentralized applications.

IV. New Frontiers for Blockchain Research

Overall, the frontiers of blockchain research are wide-spanning. We allowed for works that goes beyond Bitcoin to look at uses of blockchains to solve problems in what may appear to be wholly unrelated areas. For example, "Towards Better Availability and Accountability for IoT Updates by means of a Blockchain" presents a method for using blockchains to secure one of the riskier attack surfaces of the Internet of Things. When it actually makes sense in terms of security to deploy a blockchain to a new problem is one of the most important, if elusive, questions in blockchain research. This question is taken up seriously by the short position paper on "Oligarchic Control of Business-To-Business Blockchains," which details the advantages of having multiple roots of trust across different permissioned blockchains, as well as open research problems in permissioned blockchains.

Another emerging area where much more security and privacy research is needed is smart contracts. One of the more surprisingly successful parts of Bitcoin is its simple non-Turing complete scripting language, and a flurry of amateur programming language design has resulted from trying to work around the limitations of Bitcoin scripting, often with disastrous results in terms of security such as the DAO hack of Ethereum. The paper "Zero-Collateral Lotteries in Bitcoin and Ethereum" presents a new solution to a central problem in distributed systems as a whole: A lottery based on cryptographic commitments, carefully comparing their implementations in terms of efficiency between Bitcoin and Ethereum.

Overall, the future of privacy and security research on the blockchain looks bright: The workshop received many submissions, and only a few could be presented as mature full papers. Like the open-source community, many of the papers presented work in progress more than complete solutions or finalized products. Our domain is still young, and researchers are only dipping their toes, and many developers only starting to flirt with academic research. We therefore allowed in many short papers to add to the discussion and allow seeking for even more fruitful collaborations in the future. Still, many important subjects were unrepresented; from privacy-preserving signature aggregation to the formal verification of a new generation of smart contracts, there is much to be done. It is possible even that if there are fundamental security and privacy trade-offs related to scalability and decentralization, these could be explored using game-theoretic techniques. What was said of the price of Bitcoin in the early days will hopefully apply to the amount of high quality research in privacy and security on the blockchain we see in the future: To the moon!

Acknowledgment

The authors would like to thank Blockstream for sponsoring this workshop and Inria's work was funded by the European Unions Horizon 2020 Framework Programme for Research and Innovation (H2020-ICT-2015, ICT-10-2015) under grant agreement number 688722.

References

[1] Juan Garay, Aggelos Kiayias, and Nikos Leonardos. The bitcoin backbone protocol: Analysis and applications. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 281–310. Springer, 2015.

[2] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of computer and system sciences*, 28(2):270–299, 1984.

[3] Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M Voelker, and Stefan Savage. A fistful of bitcoins: characterizing payments among men with no names. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 127–140. ACM, 2013.

[4] Eli Ben Sasson, Alessandro Chiesa, Christina Garman, Matthew Green, Ian Miers, Eran Tromer, and Madars Virza. Zerocash: Decentralized anonymous payments from bitcoin. In *Security and Privacy (SP), 2014 IEEE Symposium on*, pages 459–474. IEEE, 2014.